Structure and Mechanism of Human DNA Polymerase η

Christian Biertümpfel[1*], Ye Zhao[2,1*], Yuji Kondo[3], Santiago Ramón-Maiques[1,6], Mark Gregory[1], Jae Young Lee[1,7], Chikahide Masutani[3], Alan R. Lehmann[5], Fumio Hanaoka[3,4] & Wei Yang[1]

[1.] Laboratory of Molecular Biology, NIDDK, NIH

[2.] Institute of Nuclear-Agricultural Sciences, Zhejiang University, China

[3.] Graduate School of Frontier Biosciences, Osaka University, Japan

[4.] Faculty of Science, Gakushuin University, Tokyo, Japan

[5.] Genome Damage and Stability Centre, University of Sussex, Brighton, UK

[6.] Current address, Spanish National Cancer Research Centre, Madrid, Spain

[7.] Current address, Dongguk University-Seoul, Seoul, Korea

[*] These two contribute equally.

Correspondence and request for material should be addressed to W.Y. or F. H.

Wei Yang
e-mail: Wei.Yang@nih.gov
LMB, NIDDK, NIH, 9000 Rockville Pike, Bldg. 5, Rm B1-03, Bethesda, MD 20892.
phone: (301) 402-4645
fax: (301) 496-0201

Fumio Hanaoka
e-mail: fumio.hanaoka@gakushuin.ac.jp
Faculty of Science, Gakushuin University, 1-5-1 Mejiro, Toshima-ku, Tokyo 171-8588, Japan
phone: +81-3-3986-0221 ex. 6457
fax: +81-3-5992-1029

Summary

The variant form of human xeroderma pigmentosum syndrome (XPV) is caused by a deficiency in DNA polymerase η (Polη) that bypasses sunlight-induced pyrimidine dimers. We report high-resolution crystal structures of human Polη at four consecutive steps during DNA synthesis through *cis-syn* cyclobutane thymine dimers. Polη acts like a molecular splint to stabilize damaged DNA in a normal B-form conformation. An enlarged active site accommodates the thymine dimer with excellent stereochemistry for two-metal ion catalysis. Two residues conserved among Polη orthologs form specific hydrogen bonds with the lesion and the incoming nucleotide to assist translesion synthesis. Based on the structures, eight Polη missense mutations causing XPV can be rationalized as undermining the "molecular splint" or perturbing the active-site alignment. The structures also shed light on the role of Polη in replicating through D loop and DNA fragile sites.

Introduction

Life on earth can scarcely escape the ultraviolet radiation (UV) of sunlight, which catalyzes covalent linkages between adjacent pyrimidines in DNA [1]. The resulting pyrimidine dimers are roadblocks to most DNA polymerases, but the human *POLH* gene, defective in the variant form of xeroderma pigmentosum (XPV), encodes a Y-family DNA polymerase, Polη, specialized for translesion synthesis (TLS) through cyclobutane pyrimidine dimers (CPDs) [2,3]. XPV is characterized by sunlight-induced pigmentation changes and a highly elevated incidence of skin malignancies. In contrast to humans, yeast lives well without Polη when chronically exposed to a low dosage of UV radiation [4]. Conversely, in humans TLS by Polη can decrease cellular sensitivity to anticancer DNA adducts, e.g. cisplatin [5]. In addition to TLS, Polη is involved in somatic hypermutation [6] and replication of D loop and DNA fragile sites [7-9].

Polη is the only one of the dozen human polymerases in which defects are unequivocally associated with cancer [10]. Among human Y-family polymerases (Rev1, Polη, ι and κ), the lesion-bypass specificity and efficiency of Polη stand out. Polη binds CPD-containing DNA better than undamaged DNA and extends primers more processively opposite and up to 2 nt beyond CPDs [11,12]. Yet Polη is completely blocked by the UV-induced pyrimidine (6-4) pyrimidone photoproducts (6-4PP) [13]. 6-4PPs occur less frequently than CPDs and distort the helical structure of DNA more severely and are thus more efficiently recognized and removed by nucleotide-excision repair proteins [14,15].

Y-family polymerases have been studied extensively in the last 14 years. Structural analyses of these TLS polymerases reveal a spacious active site that accommodates DNA adducts and non-Watson-Crick base pairs with little discrimination

[16,17]. However, the molecular basis for the specificity and efficiency of Polη in bypassing CPDs has remained a puzzle despite many crystal structures, including those of the apo and cisplatin-DNA bound yeast Polη [18,19].

We report here crystal structures of the catalytic domain of human Polη catalyzing the four-step TLS through a cis-syn cyclobutane thymine dimer. A ternary complex with a normal DNA was also determined for comparison. Two residues conserved among Polη orthologs were mutated, and the resulting defects in TLS assessed by functional studies.

Crystallization and general features of the hPolη ternary complex

The catalytic domain of human Polη (1-432aa) (abbreviated as hPolη) was expressed in *E. coli* after codon optimization (Methods). Crystals of ternary complexes were initially grown with normal DNAs, and the structure was solved and refined to 2.9Å resolution. However, a hydrophobic crystal-lattice contact between protein molecules appeared to distort hPolη-DNA interactions as compared with known structures of Y-family polymerases (Supplementary Fig. 1). Multiple amino-acid substitutions were engineered to break this lattice contact. Non-hydrolyzable dNMPNPPs were synthesized to replace dNTPs for co-crystallization of an active hPolη, DNA and Mg^{2+}. Different DNA lengths were systematically sampled and proven most effective for breaking the undesirable lattice contacts. After many trials, two new crystal forms were grown that diffracted X-rays to ~ 3Å and subsequently improved to 1.75 to 2.15Å (Supplementary Table 1-2).

The crystal structure of WT hPolη complexed with normal DNA and dAMPNPP opposite the template T (denoted Nrm) was determined and refined to 1.83Å (Methods)

(Fig. 1a). The relevance of the structure is validated by the appearance of a well-formed active site centered on two properly coordinated Mg^{2+} ions and the 3′-OH of the primer strand 3.2Å from the α-phosphate of dAMNPP poised for in-line nucleophilic attack (Fig. 1b). The single water ligand of the Mg^{2+} ions (Fig. 1b) has been observed in several DNA polymerases and may participate in catalysis [20]. Configured around the non-hydrolyzable dNTP analog, the active site is arguably the most reaction ready among the homologous A-, B- and Y-family DNA polymerases crystallized to date. The incoming nucleotide, catalytic carboxylates and metal ions are nearly superimposable with those of human Polβ (X-family) despite dissimilar tertiary structures (Supplementary Fig. 2). This active site configuration is likely conserved among all DNA polymerases and possibly all RNA polymerases as well.

Like all Y-family polymerases, hPolη contains four domains—palm, finger, thumb and little finger (LF)—with the active site in the palm domain and DNA bound between thumb and LF. Unlike the structures of yeast Polη–cisplatin complexes [19], hPolη's finger is closed and contacts the replicating base pair (Fig. 1c, Supplementary Fig. 3a), and LF interacts extensively with both template and primer in the major groove (Fig. 1d). A total of 2200$Å^2$ surface area is buried between hPolη and DNA. The finger, palm and incoming dNTP in hPolη superimpose well with those in Dpo4, Polκ and ι (Supplementary Fig. 3a). Uniquely in hPolη, however, two bases of the template strand instead of one are in the active site, and the 3′ T is base paired with the incoming dAMPNPP (Fig. 1c). Translocation of two template bases into the active site was previously observed with Dpo4, but there the 5′ base was paired with the dNTP resulting in a misalignment [21]. In hPolη a slight shift of LF relative to the finger domain enlarges

the active site, and allows it to accommodate two template bases without misalignment. Concomitantly, hPolη's thumb and LF are oriented differently relative to Dpo4, Polκ and ι (Supplementary Fig 3a). The thumb domain only contacts the DNA primer (Fig. 1d).

Structures displaying TLS through a CPD

CPD-containing oligonucleotides were synthesized to place each crosslinked thymine at the template position or 1 or 2 bps upstream (Supplementary Table 1). Four ternary complexes were crystallized and labeled as TT1, TT2, TT3 and TT4. All except TT2 were crystallized in the same space group as the complex of normal DNA (Nrm). These structures were determined by molecular replacement (Methods). After refinement, the four CPD-containing complexes are virtually superimposable with Nrm (Fig. 2a). The rmsds among proteins are 0.15 to 0.71Å over ~400 pairs of Cα atoms.

In TT1 the 3′ thymine of CPD serves as the template and forms a Watson-Crick (WC) base pair with dAMPNPP, much like an undamaged base (Fig. 2b). The 5′ thymine of CPD is turned and moved closer to the finger domain than the undamaged base in Nrm. The tighter interactions with the CPD may explain the better fidelity of hPolη in TLS (Fig. 3a). S62 forms van der Waals contacts with the 5′ base in both cases (Fig. 1c), but S62 is not conserved and replacing it with Gly increases the TLS efficiency of hPolη [22]. Even with the S62G substitution, a 6-4PP would not fit in the active site (Supplementary Fig. 4a), which explains the selectivity of Polη.

In TT2, the 5′ thymine of CPD also forms a WC base pair with dAMPNPP (Fig. 2c). Despite being crosslinked and rotationally constrained (Fig. 2a), the 3′ thymine of the CPD maintains a WC basepair with the primer, which undergoes minimal changes relative to the Nrm and TT1 complexes. This differs from CPDs complexed with a stalled

replicative polymerase, where only one crosslinked thymine is base paired [23]. In TT1 and TT2, the template thymine adopts the position most similar to that of undamaged DNA, and the crosslinked partners adjust. In TT2 the deoxyribose of the 3′ thymine moves by 2-3Å and the base is shifted toward the major groove by ~ 2Å and tilted by ~ 20° (Fig. 2c).

Each template thymine in TT1 and TT2 is hydrogen bonded to Q38 of hPolη via its O2 oxygen (Fig. 2b-c). Both cytosine and thymine are vulnerable to forming UV-induced CPDs, and both have the O2 to form the hydrogen bond with Q38 as observed in TT1 to TT4 (Fig. 2d). In Nrm Q38 is instead hydrogen bonded with deoxyribose (Fig. 2b) due to a slight shift of the DNA template (Fig. 2a). Q38 is one of the two residues uniquely conserved in the Polη family (Supplementary Fig. 5). Ala substitution of Q38 reduces the catalytic efficiency as expected (Fig. 3a). Interestingly, Q38A increases polymerase stalling after the CPD at the stage equivalent to TT3 (Fig. 3b). This stalling likely occurs because the template base cannot stack with the upstream CPD (Fig. 2e) and depends on Q38, its sole contact with the polymerase, to align with the incoming dNTP.

The second invariant residue in the Polη family is an Arg; in hPolη it is R61. In the yeast Polη-cisplatin DNA complex crystal structures, its equivalent (R73) forms cation-π interactions with the base and polar interactions with the phosphates of the incoming dNTP [19]. In the five hPolη structures, the closed finger domain prevents R61 from stacking with the base. R61 instead adopts different rotamer conformations and is hydrogen bonded with the N7 of purines (A or G) or the phosphates (Fig. 2b-d), thus favoring the anti-conformation of dNTP and preventing Hoogsteen base pairing [24]. Ala substitution of R61 reduces the polymerase efficiency as well as misincorporation of dG

opposite a template T (Fig. 3a). R61 is important for Polη to efficiently bypass cis-syn cytosine and thymine dimers but at the cost of potential misincorporation.

hPolη is a molecular splint

The DNAs in all four CPD-Polη complexes maintain a straight B-form conformation (Fig. 2a) in spite of the inability of CPDs to base stack. This unperturbed structure is a stark contrast to the bent and unwound structures of CPD-containing DNAs alone or complexed with repair proteins (Supplementary Fig. 6) [25-28]. Poor base stacking and segmentation of the DNA helices are the structural features of CPDs that are recognized by repair proteins [29]. However, when complexed with hPolη, lesion-induced perturbations are absorbed by minor adjustments to torsion-angles of the surrounding nucleotides.

hPolη acts like a molecular splint to keep the damaged DNA straight and rigid owing to a continuous and highly positively charged DNA-binding surface that interacts extensively with the four template nucleotides immediately upstream of the active site (Fig. 4a). A β strand in LF (316 to 324 aa) is nearly parallel to the template strand, and every other mainchain amide donates a hydrogen bond to the template phosphates (Fig. 1d, 4b). Each phosphate forms additional hydrogen bonds with sidechains of Arg, Lys, Tyr and Thr. In contrast, the template-binding surface in other Y-family polymerases has a gap or holes owing to separations between LF and the catalytic core (palm, thumb and finger) (Fig. 4a, Supplementary Fig. 3a). The gap is particularly large in Polκ, and the divided protein requires an N-terminal appendage (N-clasp) for stabilization [30]. Polκ and Dpo4 likely use the structural gaps (Fig. 4a) to accommodate bulky minor-groove adducts during TLS [31,32]. The gaps in Dpo4 and Dbh also promote template looping out

as a means of lesion bypass [33,34]. In hPolη LF is connected to the catalytic core by hydrogen bonds, salt bridges and hydrophobic interactions (Supplementary Fig. 3b-c). The resulting DNA-binding surface constrains the template backbone and reinforces a B-form structure in spite of CPDs.

The sidechains of R93 and R111 extend from the palm domain to the template strand and further strengthen this molecular splint. R93 connects the template strand with the incoming nucleotide via its neighbor Y92, which stacks with F18 (the steric gate) [35] and the deoxyribose of dNTP (Fig. 1c). Although this Y92/R93 pair is found in Polι and κ, only in Polη is the DNA template hydrogen bonded to R93. R111 is adjacent to the catalytic carboxylate D115 and interacts with both LF and the template strand (Fig. 5a), which explains why substitution of R111 by His leads to the XPV phenotype [36].

Dissociation of hPolη after bypassing CPDs

TLS polymerases are recruited to stalled replication forks by Rad6-Rad18 mediated ubiquitination of PCNA [37]. Dissociation following TLS, however, may be an intrinsic property of human Polη because of reduced affinity for DNA 3bp past the CPD [11,12]. In all five structures, the DNA-binding surface of hPolη makes extensive interactions with four basepairs (Fig. 4), but has reduced interactions with the template strand further upstream. Modeling a CPD 1bp further upstream from TT4 reveals steric clashes of its phosphates with hPolη in addition to the loss of favorable hydrogen bonds (Supplementary Fig. 4b).

The crosslinked thymines are shifted 1-2Å into the major groove (Fig. 2). The C5 methyl groups form van der Waals contacts with LF (L378 and F423) in TT3 and TT4 (Fig. 4c). In Nrm, the space between the DNA bases and LF is filled with water and

glycerol molecules. These major-groove interactions are lost when the CPD is 3bp beyond the active site. The position-dependent CPD-Polη interactions provide a structural basis for polymerase switching after lesion bypass.

XPV mutations

Many mutations in Polη have been identified in XPV patients, and five missense mutations R111H, A117P, T122P, G263V and R361S are within the first 432aa [36,38]. These mutations and three newly identified XPV mutations A264P, F290S and G295R (ARL, unpublished data) are modeled using the crystal structures (Fig. 5b). A117 adjoins the catalytic carboxylate E116, and substitution by Pro would cause clashes with the carbonyl oxygen of V12. T122 occurs at the beginning of an α-helix and participates in the hydrogen-bond network surrounding all three catalytic carboxylates (D13, D115 and E116) (Fig. 5c). T122P would likely perturb the active site as does A117P. R111 is a part of the molecular splint (see the previous section). R361 is hydrogen bonded with the carbonyl oxygen of P316 and anchors the β strand (316-324 aa) for template binding (Fig. 5a,d). The R111H and R361S mutations are likely to relax the protein-DNA interface and break the molecular splint.

The remaining four XPV mutations are in the thumb domain and affect primer binding, which is essential to complete the molecular splint. G263 to Val and A264 to Pro mutations would cause steric clashes with the carbonyl oxygens of L258 and K261, respectively (Fig. 5e). These clashes would disrupt primer binding by the GGK motif (259-261aa) (Fig. 1d, 4c). F290 and G295 form the hydrophobic core of the thumb domain, and F290S and G295R mutations would destabilize the structure.

Replication through D-loop and fragile sites

hPolη has been shown to extend DNA synthesis through D loop, a homologous recombination intermediate resulting from strand invasion [7,39]. Recently hPolη has also been shown to facilitate replication through fragile sites, which are associated with hairpins and other non-B-form secondary structures [40], and prevent DNA breakage [9]. Crystal lattice contacts have occasionally shed light on meaningful macromolecular interactions [41]. In hPolη the exposed W339 on the back of LF, which was identified as a culprit for distorting the hPolη ternary complexes (Supplementary Fig. 1), is stacked with the 5′ end of a neighboring primer in the $P6_1$ crystals (Nrm, TT1, TT3 and TT4) (Fig. 6a). The two symmetry-related DNA molecules are reminiscent of a D-loop structure as the template strands are nearly contiguous. W339 of hPolη is replaced by Ser in yeast Polη and by Phe, His or Lys in Polη of other eukaryotes (Supplementary Fig. 5). The lack of conservation indicates that W339 alone may not be sufficient for binding downstream DNA. In the $P2_12_12_1$ crystals (TT2), a symmetry-related DNA is packed near W339 and again two template strands are nearly contiguous (Fig. 6b). Mimics of a downstream DNA duplex have not been observed previously in polymerase-DNA co-crystals. We hypothesize that the back of hPolη may be an authentic downstream DNA-binding site, and the LF functions as a wedge to separate non-B-form DNAs. The molecular splint of hPolη, which enables it to carry out TLS through CPDs, may also allow it replicate through D loop and fragile sites.

Concluding remarks

Human Polη is specifically optimized to bypass the most ubiquitous DNA lesion. An enlarged active site accommodates CPDs, the complementary DNA-binding surface reinforces the B-form conformation, and hydrophobic residues interact with crosslinked

pyrimidines in the major groove. These features readily explain the specificity and different efficiencies of human Polη in bypassing CPDs and other lesions [42-45]. The limited interactions with the DNA minor groove may enable Polη to misincorporate nucleotides during somatic hypermutation. The high-resolution structures reported here not only provide a detailed mechanism of TLS but also generate testable models for investigating novel functions of Polη.

METHODS

The full-length human Polη gene codon-optimized for *E. coli* expression was synthesized by GenScript. The catalytic core (residues 1-432, hPolη) was cloned into modified pET28a [46], expressed in *E. coli* and purified by Ni-affinity, MonoS and Superdex75 chromatography. The His-tag was removed by PreScission protease. Mutagenesis was performed using QuikChange (Stratagene). Non-hydrolyzable dNMPNPPs were purchased from Jena Bioscience, and phosphoramidites of CPD from Glen Research. CPD oligos were synthesized and purified by TriLink Biotechnogies. Ternary complexes were prepared by mixing WT or C406M mutant hPolη and annealed DNA at a 1:1.05 molar ratio and addition of 5 mM Mg^{2+} and 1 mM non-hydrolyzable deoxynucleotides (dNMPNPP). The final protein concentration was 6-7 mg/ml. Crystals were grown in 0.1 M MES (pH 6.0), 19-21% (w/v) PEG 2K-MME and 5 mM $MgCl_2$ after several rounds of microseeding. Diffraction data were collected at sectors 22 and 23 of the APS. Phases were determined by molecular replacement [47] and multi-wavelength anomalous dispersion using selenomethionine-labeled hPolη [48]. Structures were refined using CNS [49] and interspersed with manual model building using COOT [50]. All residues are in the most favorable (97%) and allowed (2.3%) regions of Ramachandran plot except for two that are well defined by electron densities. For functional assays, the C-terminal truncated human Polη (1-511aa), which has the same TLS activity as the full-length hPolη [44], was subcloned into pET21a and readily expressed in *E. coli*. Q38A and R61A mutations were made using Mutant-K (TaKaRa BIO Inc). Steady-state kinetic assays and primer extension reactions were carried out as described [44].

References

1. Brash, D.E. Sunlight and the onset of skin cancer. *Trends Genet* 13, 410-4 (1997).
2. Masutani, C. et al. The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase η. *Nature* 399, 700-4. (1999).
3. Johnson, R.E., Kondratick, C.M., Prakash, S. & Prakash, L. hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science* 285, 263-5 (1999).
4. Hishida, T., Kubota, Y., Carr, A.M. & Iwasaki, H. RAD6-RAD18-RAD5-pathway-dependent tolerance to chronic low-dose ultraviolet light. *Nature* 457, 612-5 (2009).
5. Chaney, S.G., Campbell, S.L., Bassett, E. & Wu, Y. Recognition and processing of cisplatin- and oxaliplatin-DNA adducts. *Crit Rev Oncol Hematol* 53, 3-11 (2005).
6. Saribasak, H., Rajagopal, D., Maul, R.W. & Gearhart, P.J. Hijacked DNA repair proteins and unchained DNA polymerases. *Philos Trans R Soc Lond B Biol Sci* 364, 605-11 (2009).
7. McIlwraith, M.J. et al. Human DNA Polymerase η Promotes DNA Synthesis from Strand Invasion Intermediates (D-loops) of Homologous Recombination. *Mol Cell* 20, 783-792. (2005).
8. Bugreev, D.V., Hanaoka, F. & Mazin, A.V. Rad54 dissociates homologous recombination intermediates by branch migration. *Nat Struct Mol Biol* 14, 746-53 (2007).
9. Rey, L. et al. Human DNA polymerase eta is required for common fragile site stability during unperturbed DNA replication. *Mol Cell Biol* 29, 3344-54 (2009).
10. Di Lucca, J. et al. Variants of the xeroderma pigmentosum variant gene (POLH) are associated with melanoma risk. *Eur J Cancer* 45, 3228-36 (2009).
11. Kusumoto, R., Masutani, C., Shimmyo, S., Iwai, S. & Hanaoka, F. DNA binding properties of human DNA polymerase eta: implications for fidelity and polymerase switching of translesion synthesis. *Genes Cells* 9, 1139-50 (2004).
12. McCulloch, S.D. et al. Preferential *cis-syn* thymine dimer bypass by DNA polymerase η occurs with biased fidelity. *Nature* 428, 97-100 (2004).
13. Yao, J., Dixon, K. & Carty, M.P. A single (6-4) photoproduct inhibits plasmid DNA replication in xeroderma pigmentosum variant cell extracts. *Environ Mol Mutagen* 38, 19-29 (2001).
14. Scrima, A. et al. Structural basis of UV DNA-damage recognition by the DDB1-DDB2 complex. *Cell* 135, 1213-23 (2008).
15. Sugasawa, K. XPC: its product and biological roles. *Adv Exp Med Biol* 637, 47-56 (2008).
16. Yang, W. & Woodgate, R. What a difference a decade makes: insights into translesion DNA synthesis. *Proc Natl Acad Sci U S A* 104, 15591-8 (2007).
17. Broyde, S., Wang, L., Rechkoblit, O., Geacintov, N.E. & Patel, D.J. Lesion processing: high-fidelity versus lesion-bypass DNA polymerases. *Trends Biochem Sci* 33, 209-19 (2008).
18. Trincao, J. et al. Structure of the catalytic core of *S. cerevisiae* DNA polymerase η: implications for translesion DNA synthesis. *Mol Cell* 8, 417-26. (2001).

19. Alt, A. et al. Bypass of DNA lesions generated during anticancer treatment with cisplatin by DNA polymerase eta. *Science* 318, 967-70 (2007).
20. Wang, L., Broyde, S. & Zhang, Y. Polymerase-tailored variations in the water-mediated and substrate-assisted mechanism for nucleotidyl transfer: insights from a study of T7 DNA polymerase. *J Mol Biol* 389, 787-96 (2009).
21. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* 107, 91-102. (2001).
22. Glick, E., Vigna, K.L. & Loeb, L.A. Mutations in human DNA polymerase η motif II alter bypass of DNA lesions. *Embo J* 20, 7303-12 (2001).
23. Li, Y. et al. Nucleotide insertion opposite a cis-syn thymine dimer by a replicative DNA polymerase from bacteriophage T7. *Nat Struct Mol Biol* 11, 784-90 (2004).
24. Ling, H., Boudsocq, F., Plosky, B.S., Woodgate, R. & Yang, W. Replication of a cis-syn thymine dimer at atomic resolution. *Nature* 424, 1083-7 (2003).
25. Park, H. et al. Crystal structure of a DNA decamer containing a cis-syn thymine dimer. *Proc Natl Acad Sci U S A* 99, 15965-70 (2002).
26. Vassylyev, D.G. et al. Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* 83, 773-82 (1995).
27. Min, J.H. & Pavletich, N.P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* 449, 570-5 (2007).
28. Mees, A. et al. Crystal structure of a photolyase bound to a CPD-like DNA lesion after in situ repair. *Science* 306, 1789-93 (2004).
29. Yang, W. Poor base stacking at DNA lesions may initiate recognition by many repair proteins. *DNA Repair (Amst)* 5, 654-66 (2006).
30. Lone, S. et al. Human DNA polymerase kappa encircles DNA: implications for mismatch extension and lesion bypass. *Mol Cell* 25, 601-14 (2007).
31. Bauer, J. et al. A structural gap in Dpo4 supports mutagenic bypass of a major benzo[a]pyrene dG adduct in DNA through template misalignment. *Proc Natl Acad Sci U S A* 104, 14905-10 (2007).
32. Jia, L., Geacintov, N.E. & Broyde, S. The N-clasp of human DNA polymerase kappa promotes blockage or error-free bypass of adenine- or guanine-benzo[a]pyrenyl lesions. *Nucleic Acids Res* 36, 6571-84 (2008).
33. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Snapshots of replication through an abasic lesion; structural basis for base substitutions and frameshifts. *Mol Cell* 13, 751-62 (2004).
34. Wilson, R.C. & Pata, J.D. Structural insights into the generation of single-base deletions by the Y family DNA polymerase dbh. *Mol Cell* 29, 767-79 (2008).
35. Jarosz, D.F., Godoy, V.G., Delaney, J.C., Essigmann, J.M. & Walker, G.C. A single amino acid governs enhanced activity of DinB DNA polymerases on damaged templates. *Nature* 439, 225-8 (2006).
36. Broughton, B.C. et al. Molecular analysis of mutations in DNA polymerase eta in xeroderma pigmentosum-variant patients. *Proc Natl Acad Sci U S A* 99, 815-20 (2002).
37. Lehmann, A.R. et al. Translesion synthesis: Y-family polymerases and the polymerase switch. *DNA Repair (Amst)* 6, 891-9 (2007).

38. Tanioka, M. et al. Molecular analysis of DNA polymerase eta gene in Japanese patients diagnosed as xeroderma pigmentosum variant type. *J Invest Dermatol* 127, 1745-51 (2007).
39. Kawamoto, T. et al. Dual roles for DNA polymerase eta in homologous DNA recombination and translesion DNA synthesis. *Mol Cell* 20, 793-9 (2005).
40. Durkin, S.G. & Glover, T.W. Chromosome fragile sites. *Annu Rev Genet* 41, 169-92 (2007).
41. Zhang, X., Gureasko, J., Shen, K., Cole, P.A. & Kuriyan, J. An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* 125, 1137-49 (2006).
42. Bassett, E. et al. Efficiency of extension of mismatched primer termini across from cisplatin and oxaliplatin adducts by human DNA polymerases beta and eta in vitro. *Biochemistry* 42, 14197-206 (2003).
43. Kokoska, R.J., McCulloch, S.D. & Kunkel, T.A. The efficiency and specificity of apurinic/apyrimidinic site bypass by human DNA polymerase eta and Sulfolobus solfataricus Dpo4. *J Biol Chem* 278, 50537-45 (2003).
44. Kusumoto, R., Masutani, C., Iwai, S. & Hanaoka, F. Translesion synthesis by human DNA polymerase eta across thymine glycol lesions. *Biochemistry* 41, 6090-9 (2002).
45. Masutani, C., Kusumoto, R., Iwai, S. & Hanaoka, F. Mechanisms of accurate translesion synthesis by human DNA polymerase η. *EMBO J* 19, 3100-3109 (2000).
46. Wang, F. & Yang, W. Structural insight into translesion synthesis by DNA Pol II. *Cell* 139, 1279-89 (2009).
47. McCoy, A.J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63, 32-41 (2007).
48. Hendrickson, W.A., Horton, J.R. & LeMaster, D.M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* 9, 1665-72 (1990).
49. Brünger, A.T. et al. Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.* D54, 905-921 (1998).
50. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and Development of Coot. *Acta Crystallographica Section D* 66, 486-501 (2010).

Accession numbers

The coordinates and structure factors have been deposited to Protein Data Bank with accession codes XXX1 (Nrm), XXX2 (TT1), XXX3 (TT2), XXX4 (TT3) and XXX5 (TT4)

Acknowledgment

We thank D. Leahy, M. Gellert and R. Craigie for critical reading of the manuscript. The research was funded by the intramural research program of NIDDK, NIH and grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. Y.Z. is a recipient of Chinese Ministry of Education scholarship and joint PhD student in NIH-Zhejiang University Graduate Partnership Program. S.R-M received a fellowship from the Human Frontiers Science Program.

C.B. determined the five structures; Y.Z. grew the crystals; Y.K. did the kinetic and bypass assays; S.R.-M. determined the type 1 structure; M.G. prepared the clone and type 1 crystals; J.-Y.L. made mutants; C.M. designed the functional assays; A.R.L. identified the unpublished XPV mutations; F.H. conceived the project; and W.Y. supervised the structure determination. C.B., Y.Z., F.H. and W.Y. prepared the manuscript.

Figure legends

Figure 1 Structure of hPolη. **a.** The ternary complex of hPolη with a normal DNA. Protein domains are shown in distinct colors and labeled. The DNA template is colored orange and the primer yellow. Oxygen and nitrogen atoms are colored red and blue, respectively. dAMPNPP is shown as stick-and-balls, and Mg^{2+} as purple spheres. All structural figures were made using PyMOL (www.pymol.org). **b.** The active site. Mg^{2+} coordination is indicated by pale yellow dashed lines. The 3′-OH of the primer strand is 3.2Å from the α-phosphate as indicated by the red dashed line. **c.** hPolη-DNA interactions around the active site. Protein side chains from the finger and palm domain are shown as light blue and pink sticks, respectively. **d.** Interactions with the upstream DNA. LF (shown as light purple ribbon diagram) contacts both template and primer, and the thumb domain (shown in green) only makes 3-4 hydrogen bonds with the primer strand. Side chains that make DNA contacts are highlighted in sticks and labeled.

Figure 2 Structures of lesion DNAs. **a.** Overlay of TT1, TT2, TT3 and TT4 DNA with the DNA in Nrm (undamaged). CPDs are shown as red sticks. DNA and Mg^{2+} are colored as in Fig. 1. **b-d**. The replicating base pairs in four ternary complexes (highly similar in TT3 and TT4) are each shown with hPolη residues that contact the bases and deoxyribose. Hydrogen bonds are shown as dashed lines. **e.** CPD in TT3 and TT4 forms WC-like hydrogen bonds but cannot base stack with neighboring nucleotides.

Figure 3 Functional analyses of Q38A and R61A Polη. **a.** Efficiencies (k_{cat}/K_M) of WT and two mutants in correct and incorrect nucleotide incorporation on normal (ND) and CPD DNA. These values are derived from Supplementary Table 3. **b.** CPD-bypass by WT (left) and Q38A mutant Polη (right). The DNA sequence for primer extension assays

is shown. Magenta colored TT represents the CPD. The same DNA template with a 2nt longer primer was used for the kinetic assay shown in Fig. 3a. Reactions were carried out at a substrate:enzyme molar ratio of 100:1 for the indicated incubation time. Termination probabilities ($[I_n] / \Sigma_{n\text{-}end}[I]$, where I_n is band intensity at the n^{th} position) on normal (ND) and CPD DNA are plotted below.

Figure 4 hPolη is a molecular splint. **a.** The DNA binding surfaces of various Y-family polymerases are shown with electrostatic potentials. Blue and red represent the positive and negative charge potential at the + and - 10 kT/e scale, respectively. The thumb domain in all cases and the N-clasp of Polκ are removed for clarity. Five phosphorus atoms at position -1 to -5 in the hPolη Nrm (yellow) and TT4 (orange) structures are also shown. **b.** The extensive interactions between hPolη and the template strand in TT4 (as an example) are shown. Hydrogen bonds are indicated by dashed lines. The finger and LF domains are colored in light purple and light green. Two residues forming hydrophobic interactions with the CPD bases are shown in teal. **c.** A view approximately 90° from b. Detailed hPolη-DNA interactions are shown in Supplementary Fig. 7.

Figure 5 XPV mutations. **a.** R111 contacts DNA template and stacks with P316 of LF. **b.** Mapping of 9 missense XPV mutations. The protein is represented by the Cα trace, DNA as tube-and-ladders, and the altered residues are shown as cyan ball-and-sticks. **c.** Local interactions of A117 and T122 and their relation to the catalytic carboxylates. Most side chains are removed for clarity. **d.** R361 supports P316 and the β strand that forms an extensive hydrogen-bond network with the DNA template. **e.** G263 and A264 are adjacent to the thumb-DNA primer interface. G263V and A264P substitutions are

modeled and would clash with L258 and K261. Electron densities are shown as meshes in panel c and e.

Figure 6 Potential interactions with downstream DNA. **a.** A front view of the symmetry-related DNA molecules in the $P6_1$ crystal. hPolη is shown in blue-grey cartoon with semi-transparent molecular surface. DNAs are colored in orange (template) and yellow (primer). The 5′ end of the symmetry-related primer strand stacks with W339 (shown in magenta). **b.** The back view of the complex. The symmetry-related DNA in the $P2_12_12_1$ crystal is also included and colored in darker shades. It may represent a second location of the downstream duplex or mimic additional DNA hairpin structures at a fragile site. R81 and R84 near the symmetry-related DNA are highlighted in blue.

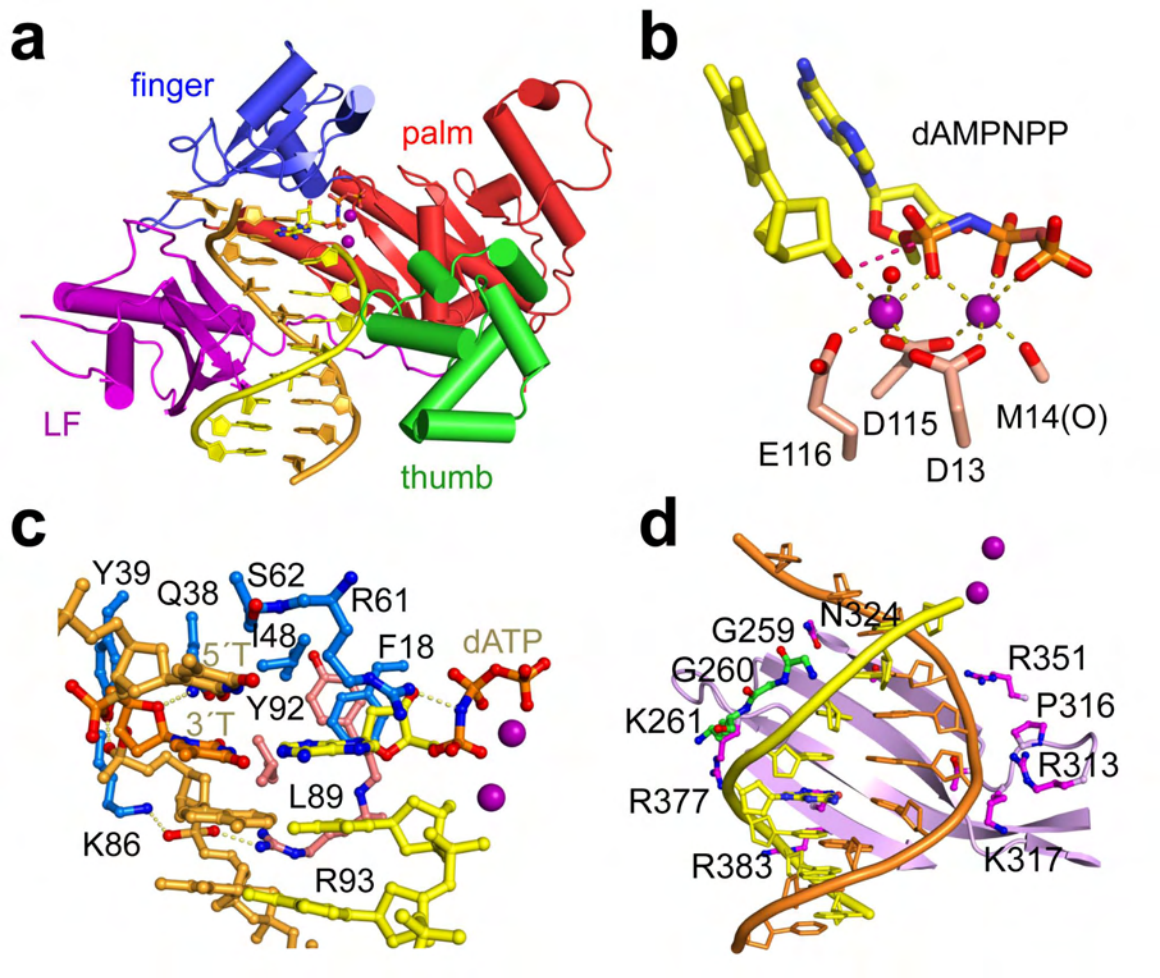
a
finger
palm
LF
thumb
b
dAMPNPP
E116
D115
M14(O)
D13
c
Y39
Q38
S62
R61
I48
5′T
F18
dATP
3′T
Y92
L89
K86
R93
d
N324
G259
G260
R351
P316
K261
R313
R377
R383
K317

Fig. 1

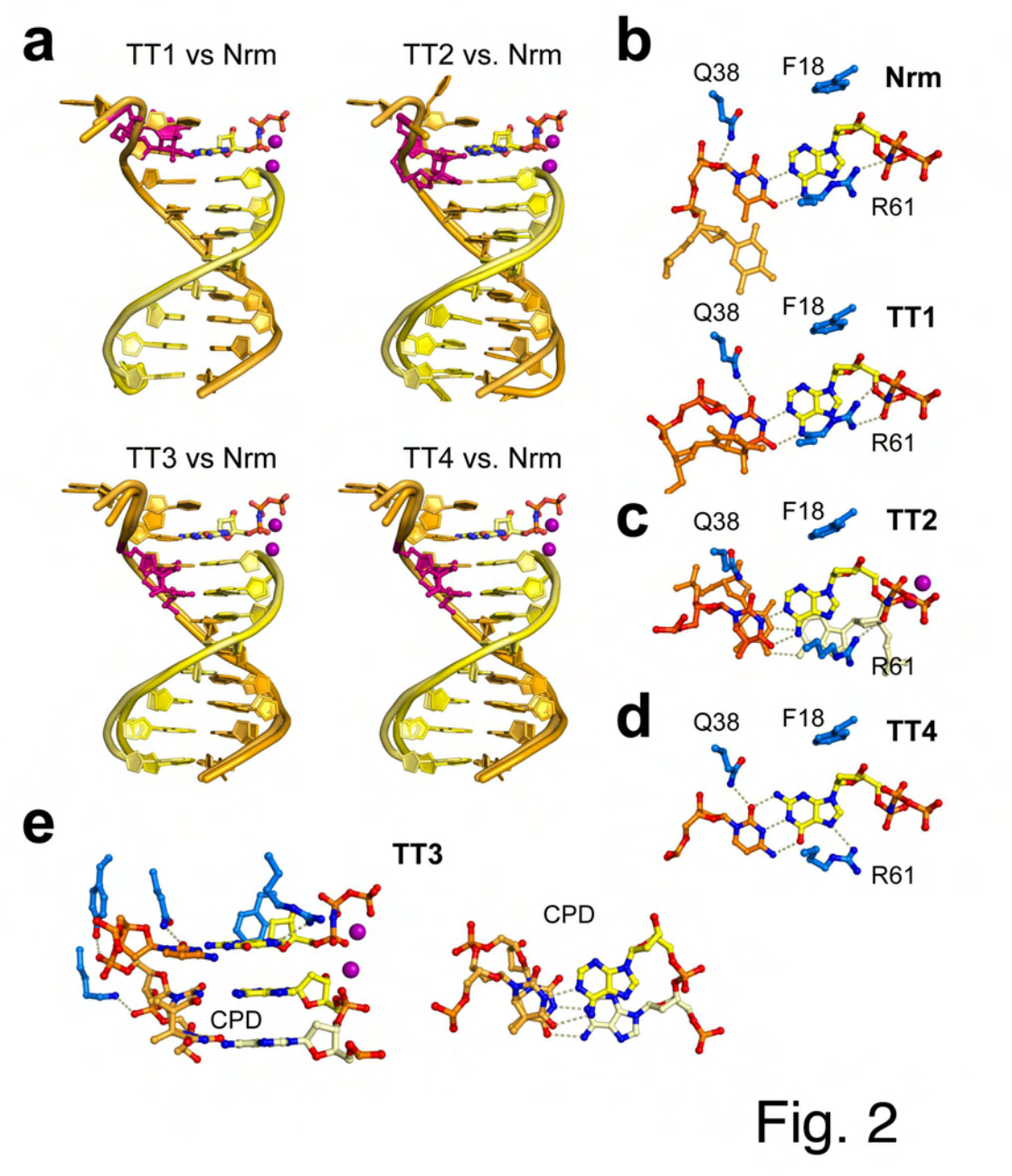
a
TT1 vs Nrm
TT2 vs. Nrm
TT3 vs Nrm
TT4 vs. Nrm
b
Q38
F18
Nrm
R61
Q38
F18
TT1
R61
c
Q38
F18
TT2
R61
d
Q38
F18
TT4
R61
e
TT3
CPD
CPD

Fig. 2

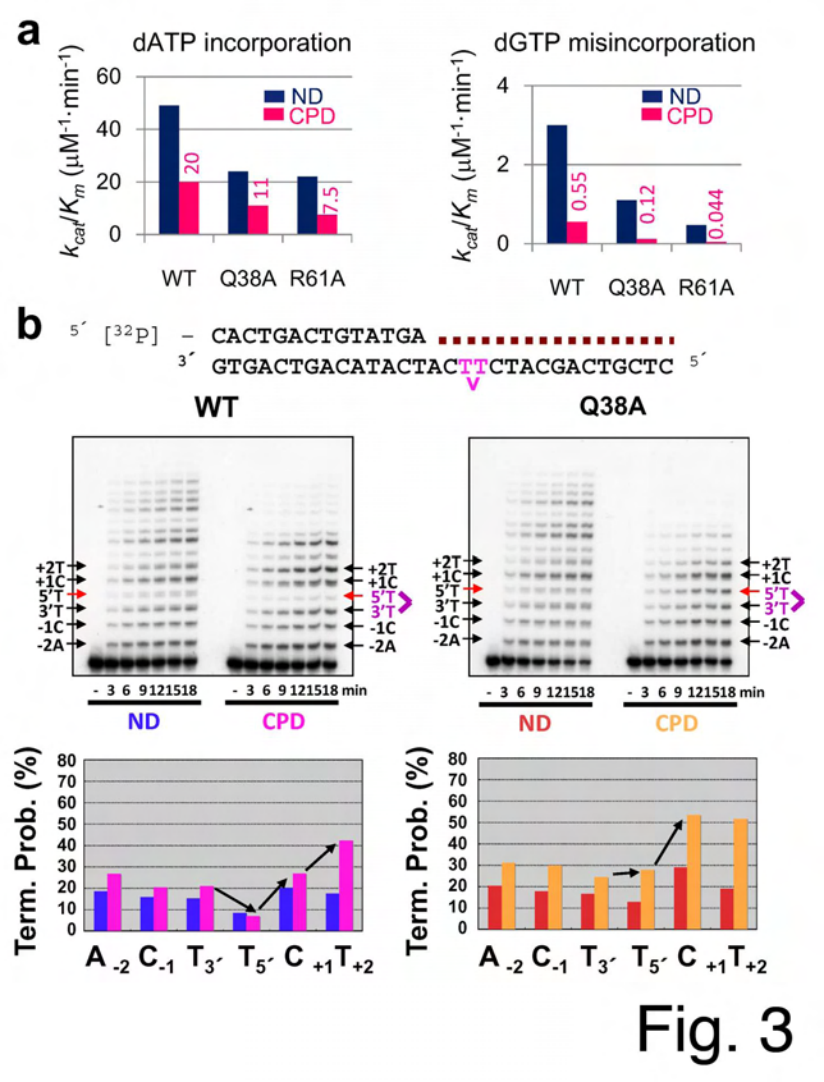
a
dATP incorporation
k_{cat}/K_m (μM^{-1}·min^{-1})
60
40
20
0
ND
CPD
20
11
7.5
WT
Q38A
R61A
dGTP misincorporation
k_{cat}/K_m (μM^{-1}·min^{-1})
4
2
0
ND
CPD
0.55
0.12
0.044
WT
Q38A
R61A
b
5′ [32P] – CACTGACTGTATGA
3′ GTGACTGACATACTACTTCTACGACTGCTC 5′
V
WT
Q38A
+2T
+1C
5′T
3′T
-1C
-2A
- 3 6 9 12 15 18 min
ND
CPD
Term. Prob. (%)
80
70
60
50
40
30
20
10
0
A -2 C -1 T 3′ T 5′ C +1 T +2

Fig. 3

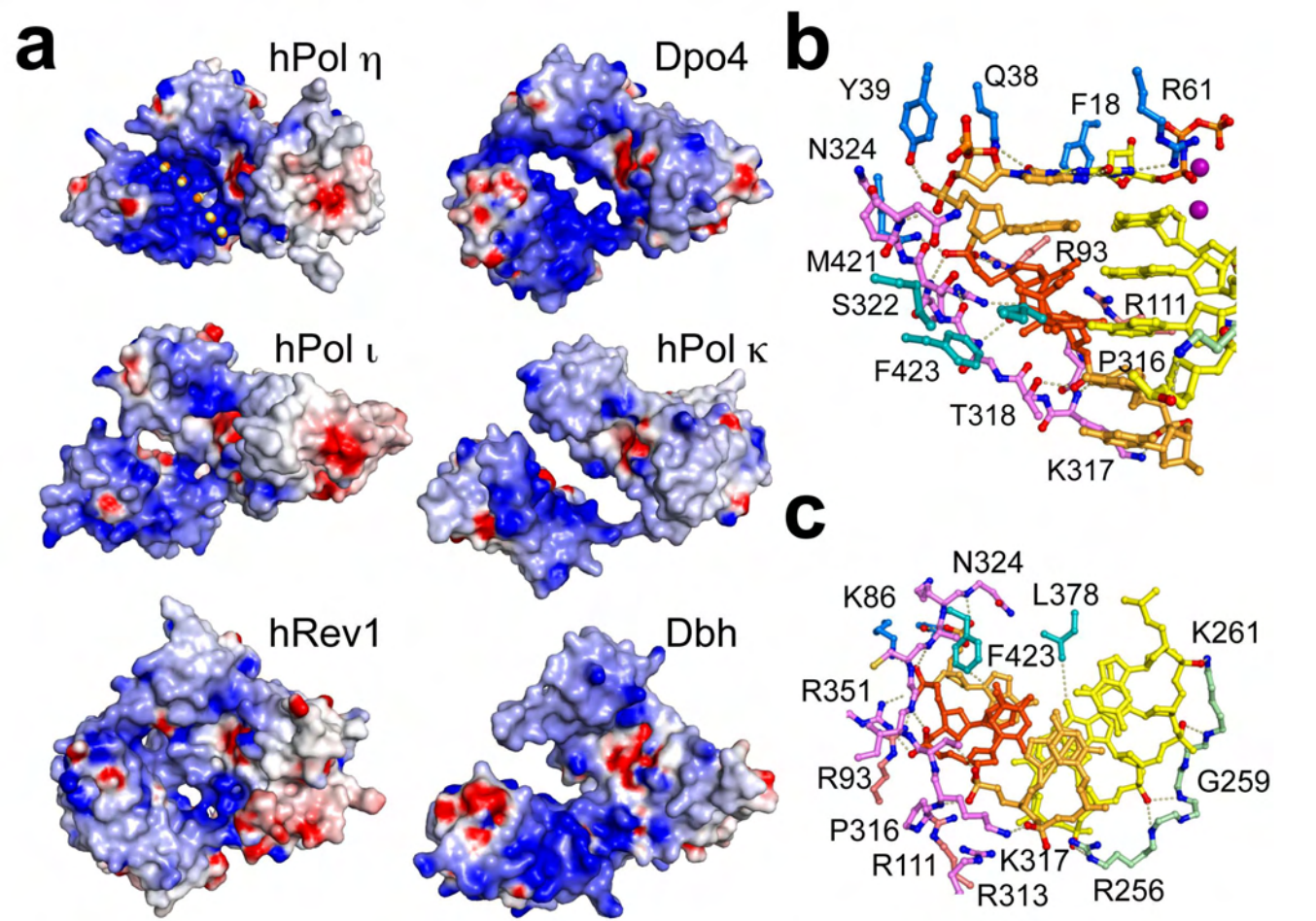
a
hPol η
Dpo4
hPol ι
hPol κ
hRev1
Dbh
b
Y39
Q38
F18
R61
N324
M421
R93
S322
R111
F423
P316
T318
K317
c
N324
K86
L378
K261
F423
R351
R93
G259
P316
R111
K317
R313
R256

Fig. 4

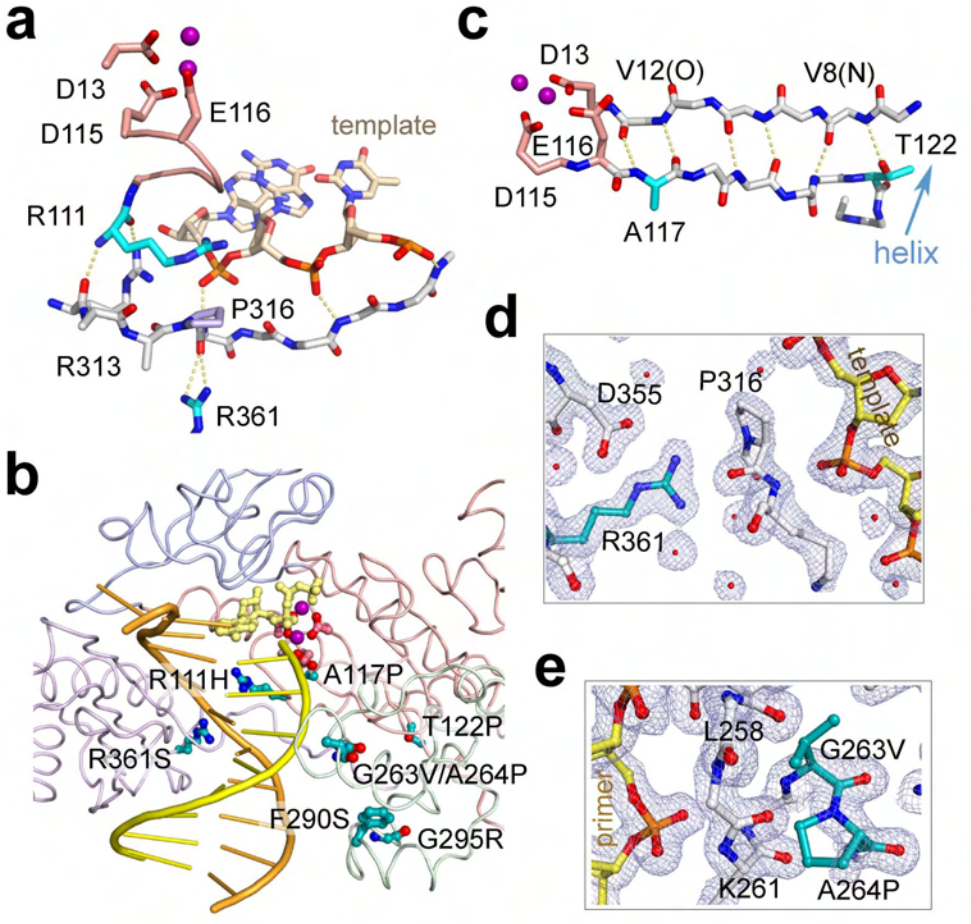
a
D13
D115
E116
template
R111
P316
R313
R361
b
R111H
A117P
T122P
R361S
G263V/A264P
F290S
G295R
c
D13
V12(O)
V8(N)
E116
T122
D115
A117
helix
d
D355
P316
template
R361
e
L258
G263V
primer
K261
A264P

Fig. 5

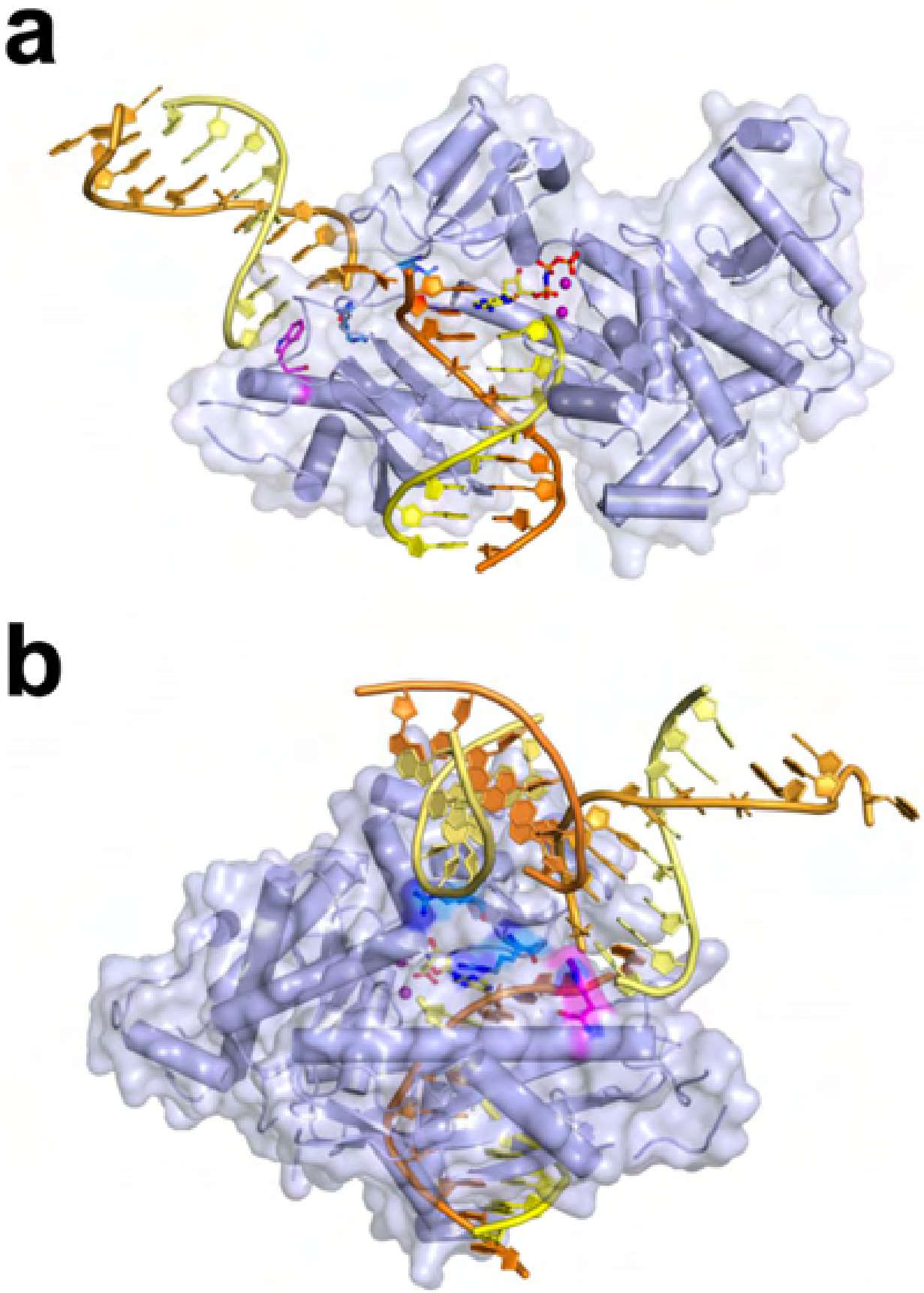

Fig. 6